Article

Synergistic Use of Multi-Temporal RADARSAT-2 and VENµS Data for Crop Classification Based on 1D Convolutional Neural Network

Chunhua Liao [1], **Jinfei Wang** [1,*], **Qinghua Xie** [2], **Ayman Al Baz** [1], **Xiaodong Huang** [3], **Jiali Shang** [4] **and Yongjun He** [1]

[1] Department of Geography, The University of Western Ontario, London, ON N6A 5C2, Canada; cliao33@uwo.ca (C.L.); aalbaz2@uwo.ca (A.A.B.); yhe563@uwo.ca (Y.H.)

[2] School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China; xieqh@cug.edu.cn

[3] Applied Geosolutions, Durham, NH 03824, USA; xhuang@appliedgeosolutions.com

[4] Research Branch, Agriculture and Agri-Food Canada, Ottawa, ON K1A 0C6, Canada; jiali.shang@agr.gc.ca

* Correspondence: jfwang@uwo.ca

Abstract: Annual crop inventory information is important for many agriculture applications and government statistics. The synergistic use of multi-temporal polarimetric synthetic aperture radar (SAR) and available multispectral remote sensing data can reduce the temporal gaps and provide the spectral and polarimetric information of the crops, which is effective for crop classification in areas with frequent cloud interference. The main objectives of this study are to develop a deep learning model to map agricultural areas using multi-temporal full polarimetric SAR and multi-spectral remote sensing data, and to evaluate the influence of different input features on the performance of deep learning methods in crop classification. In this study, a one-dimensional convolutional neural network (Conv1D) was proposed and tested on multi-temporal RADARSAT-2 and VENµS data for crop classification. Compared with the Multi-Layer Perceptron (MLP), Recurrent Neural Network (RNN) and non-deep learning methods including XGBoost, Random Forest (RF), and Support Vector Machina (SVM), the Conv1D performed the best when the multi-temporal RADARSAT-2 data (Pauli decomposition or coherency matrix) and VENµS multispectral data were fused by the Minimum Noise Fraction (MNF) transformation. The Pauli decomposition and coherency matrix gave similar overall accuracy (OA) for Conv1D when fused with the VENµS data by the MNF transformation (OA = 96.65 ± 1.03% and 96.72 ± 0.77%). The MNF transformation improved the OA and F-score for most classes when Conv1D was used. The results reveal that the coherency matrix has a great potential in crop classification and the MNF transformation of multi-temporal RADARSAT-2 and VENµS data can enhance the performance of Conv1D.

Keywords: crop classification; RADARSAT-2; VENµS; data fusion; deep learning; convolutional neural network

1. Introduction

Annual crop inventory information is important for many agriculture applications and government statistics. Remote sensing satellite imagery has provided an efficient means for crop classification. Traditionally, optical data have been widely used by providing spatial and spectral information of land covers. For crop classification, remotely sensed time series data have been proven beneficial due to different temporal features associated with different crop types. However, due to weather condition, continuous optical time series data may be difficult to acquire. Polarimetric synthetic aperture radar

(SAR) time series data can provide not only structural information but also continuous temporal changes of crops due to its capability of penetrating clouds and light rains. Therefore, multi-temporal polarimetric SAR data have been adopted for crop classifications [1–4]. The availability of various sources of satellite imagery enables to provide spatial, temporal, spectral and even structural features of land covers. It has been reported that the integration of optical and SAR data is able to reduce temporal gaps [5] and provide both spectral and structural features of land covers, which is beneficial for crop classification [6]. Previous studies have shown that the synergistic use of polarimetric SAR and optical data can increase classification accuracy in cropland areas [5,7,8].

A common method for the synergistic use of multi-source remote sensing data in land cover classification is data fusion, a process of combining images obtained by different sensors to form a composite image. Data fusion mainly focuses on the improvement of spatial resolution, structural and textural details [9]. Most of the studies related to data fusion are at pixel level [10]. The three most commonly used Optical-Radar fusion methods at pixel level are Principal Component Analysis (PCA) [11], intensity-hue-saturation (IHS) [12], and discrete wavelet transform [13]. According to previous studies, PCA is the most preferred method of the three methods [14], while the Minimum Noise Fraction (MNF) transformation performs two separate standard PCA transformation of the noise-whitened data [15], and also aims to produce principal components by maximizing the signal-to-noise ratio of the data [16].

Traditionally, supervised machine learning classification approaches such as Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Neural Networks (NN) and Decision Tree (DT) were adopted for crop classification using the integration of optical and SAR data. In recent years, deep learning (DL) has drawn attention in the remote sensing community due to the large amount of data sources and improved hardware resources. Convolutional neural network (CNN) and Recurrent Neural Network (RNN) are two deep learning architectures that have been successfully applied to remote sensing data in crop classification [9,17]. RNNs are designed for sequential or temporal data analysis such as signal processing, natural language processing and speech recognition [11], and they have shown success in remote sensing time series applications [9,17–20]. Convolutional neural networks (CNNs) have been widely used in various remote sensing applications such as land cover mapping [21], change detection [22], building extraction [23]. CNNs include one-dimensional CNNs (1D-CNNs), two dimensional CNNs (2D-CNNs) and three-dimensional CNNs (3D-CNNs). 1D-CNNs are usually applied to pixel-based hyperspectral or multi-temporal remote sensing data [9]. 2D-CNNs are generally adopted to extract features in spatial dimension such as object detection [24] and semantic segmentation [25]. 3D-CNNs consider both spatial and temporal/spectral dimension. The 1D-, 2D-, and 3D-CNNs have been applied to remote sensing images in cropland classification [9,26,27]. A few studies have demonstrated that the CNNs are superior to RNNs in crop classification using time series data [9,18,28]. In this study, the crop lands can be regarded as homogenous area, we focus on 1D-CNN at pixel level.

Due to the free access, Sentinel-1 SAR time series data are the most used dataset for crop classification using deep learning methods in recent years. However, Sentinel-1 data have only two polarizations (VV + VH). When RADARSAT-2 full polarimetric SAR data were used in cropland classification, the polarimetric SAR parameters were usually extracted from the coherency matrix using different decomposition methods such as Pauli decomposition, Cloude–Pottier decomposition, Freeman–Durden decomposition [29], Neumann decomposition [3], and the optimum power [30] for crop classification. It has been studied that the elements of the coherency matrix of the fully polarimetric SAR data also show good performance in crop classification since the coherency matrix is the basic matrix representing the information of the polarimetric SAR data [15].

To our best knowledge, no study has been tested on deep learning methods using the combination of multi-temporal RADARSAT-2 fully polarimetric SAR and optical data for cropland classification. The main objectives of this study are (1) to develop a deep learning model to map cropland areas using

multi-temporal full polarimetric SAR and multi-spectral remote sensing data, and (2) to evaluate the influence of different input features on the performance of deep learning methods in crop classification.

2. Materials

2.1. Study Site

The study site is located in the agricultural area in Mixedwood Plains Ecozone Southwestern Ontario (Figure 1), characterized by an abundant water supply and productive soils for agriculture. The dominant crops in the study site are winter wheat, corn, soybeans, and forage, including alfalfa and grass. The crop types in each field changed every year due to crop rotation. Generally, corn and soybean are seeded in May and harvested in October. Winter wheat in this study site is seeded in October the previous year and harvested in July of the following year.
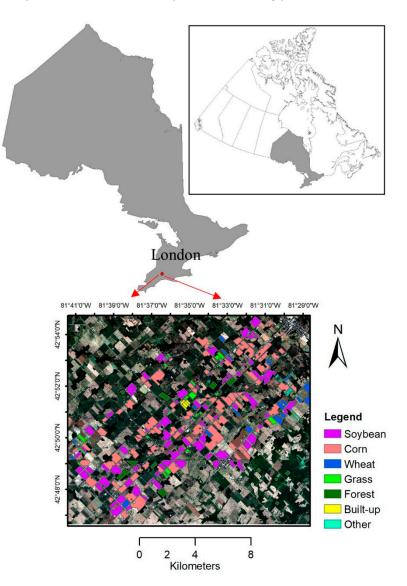

Figure 1. Location of the study site, RGB image of the VENµS data acquired in July and ground truth data.

2.2. Ground Truth Data

From July to October 2018, intensive field surveys were conducted nearly every week. Field data including crop types, crop phenology, crop height, leaf area index (LAI), and soil moisture were

collected. In addition, a general landcover survey was also conducted in October. Soybean, corn, winter wheat, alfalfa, grass, tobacco and squash were the crop types found in the field survey in this study site. Forest and built-up classes were defined in Google Map. The ground truth data were digitized into polygons (Figure 1). We avoided including field boundaries when digitizing the ground truth data to ensure that there is no mixed pixel. Alfalfa and grass were aggregated into a single forage class. Tobacco and squash were aggregated into other class due to their limited sample sizes. Therefore, seven classes were defined to represent the land cover in the study site. Then the polygons were converted to raster with 10 m spatial resolution in order to be consistent with the pixel size of the processed remote sensing data. The ground truth data were split into training, validation and testing datasets. The training and testing sets were used to train and test individual classification algorithms. The validation set was used to select the optimal hyper-parameters of the two deep learning methods. As the pixels in the same field are homogenous and highly correlated [18], we need to avoid the three datasets being from the same field. The ground truth data were split randomly into five mutually exclusive folds at polygon level, and then the same number of pixels were selected for each subset as the number of pixels in each polygon was not equal. Then one fold (20%) of the datasets was used as validation data and another fold (20%) was used as testing data. And the remaining three folds (60%) of the datasets were used as training data. This combination was repeated five times. Hence, each algorithm was evaluated based on five different train/test splits. Table 1 shows the number of fields and the pixels used for training, validation and testing datasets.

Table 1. Sample size of training, validation and testing datasets.

Land Cover	Field Polygons	Training Pixels	Validation Pixels	Testing Pixels
Soybean (Purple)	80	60,000	20,000	20,000
Corn (Coral)	104	66,000	22,000	22,000
Winter Wheat (Blue)	20	11,400	3800	3800
Forage (Green)	29	9000	3000	3000
Forest (Dark Green)	13	7500	2500	2500
Built-up (Yellow)	5	1440	480	480
Other (Cyan)	10	1110	370	370

2.3. Remote Sensing Data

2.3.1. RADARSAT-2 Data

A total of 10 fine-quad wide beam mode (FQW) RADARSAT-2 polarimetric SAR images were acquired throughout the 2018 growing season from July to October (Table 2). The RADARSAT-2 data are single look complex (SLC) format containing four polarizations HH, HV, VV, and VH. The revisit time for the same beam mode of RADARSAT-2 data is 24 days.

Table 2. Acquisition dates of RADARSAT-2 data and their corresponding incidence angles.

Date	Beam Mode	Incidence Angle	Orbit
2018-07-01	FQ10W	28.4–31.6	ASC
2018-07-08	FQ5W	22.5–26.0	ASC
2018-07-25	FQ10W	28.4–31.6	ASC
2018-08-01	FQ5W	22.5–26.0	ASC
2018-08-18	FQ10W	28.4–31.6	ASC
2018-08-25	FQ5W	22.5–26.0	ASC
2018-09-01	FQ1W	17.5–21.2	ASC
2018-09-08	FQ4W	21.3–24.8	DES
2018-09-15	FQ9W	27.2–30.5	DES
2018-10-05	FQ10W	28.4–31.6	ASC

The coherency matrices (T3) were extracted from the RADARSAT-2 data, and a 9 × 9 Boxcar filter was applied to suppress the inherent speckle noise. This window size is selected to preserve the sufficient Equivalent Numbers of Looks (ENL) and to keep details as many as possible. A Digital Elevation Model (DEM) of Ontario, Canada with resolution of 30 m was used for geocoding, and the output spatial resolution is 10 m in UTM geographical coordinate system. The linear polarizations, Pauli decomposition, Freeman-Durden decomposition, and Cloude-Pottier decomposition were conducted based on each geocoded coherency matrix, respectively. Then the overlapping area of multi-temporal RADARSAT-2 images that covers the study site was selected, and all the RADASART-2 data were resized.

2.3.2. VENµS Data

Vegetation and Environment monitoring on a New Micro-Satellite (VENµS) was launched in August 2017. It is a near polar sun-synchronous orbit microsatellite developed jointly by the Israel Space Agency (ISA) and the French space agency (CNES). It provides images with 12 narrow spectral bands ranging from 420 nm to 910 nm at high spatial and temporal resolutions (5–10 m every 2 days). As our study site has been selected as one of the study areas in the world, the data can be downloaded from Theia Data Center (https://www.theia-land.fr/en/data-and-services-for-the-land/) for free. In this study, two cloud-free VENµS level 2 (L2A) surface reflectance products with 10 m spatial resolution acquired on 11 June 2018 and 9 July 2018 were utilized to fill the temporal gap and provide the spectral information. The reflectance data were divided by 1000 so that the range was between 0 and 1. The range of polarimetric SAR backscattering was also between 0 and 1. Thus, the scale of the optical and SAR data was normalized.

2.4. Data Preparation

Firstly, the coherency matrix, backscattering coefficients at linear polarization and polarimetric features from three polarimetric decompositions (Pauli decomposition, Cloude-Pottier decomposition, Freeman-Durden decomposition) of the multi-temporal RADARSAT-2 data were stacked respectively and tested as inputs for all the classifiers. Then, the polarimetric SAR parameters were combined with the VENµS multispectral data, respectively, and the MNF transformation was conducted on the combination of the two sources of data. As the values of the backscattering coefficients and spectral reflectance both range between 0 and 1, we did not apply any other normalization approach to the original features.

Labels were created from the training, validation and testing ground truth datasets. The input features were split into training, validation and testing features according to the label datasets. They were shuffled in order to reduce variance and overfitting.

3. Methods

In this study, a one-dimensional convolutional neural network (Conv1D) was proposed for cropland classification. For comparison purpose, a multi-layer perceptron (MLP) and a recurrent neural network (RNN) were built and tested as deep learning methods, and the XGBoost, RF and SVM were testes as non-deep learning benchmark classifiers. The hyperparameters of all the classifiers were trained and optimized using the MNF transformation of multi-temporal RADASART-2 coherency matrix and VENµS multispectral data. Then, different scenarios of remote sensing datasets and the ground truth data were used to train and validate the optimized classifiers based on a five-fold cross validation process. The final accuracy assessment was based on the average and standard deviation of the five folds cross validation. A flowchart of the methodology used in this study is presented in Figure 2. Basically, there are four steps. (1) Data acquisition and preprocessing; (2) training of hyperparameters for the deep learning methods; (3) training and cross validation of the optimized classifier using remote sensing different datasets and ground truth data; (4) final classification map generation and accuracy assessment using the trained classifiers.
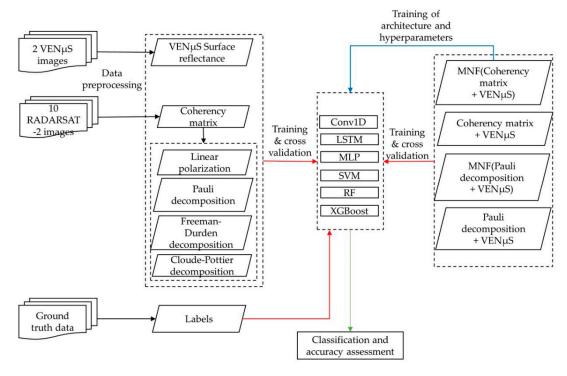
Figure 2. Flowchart of the methodology. The black arrow denotes data processing; the blue indicates means training of architecture and hyperparameters; the red arrow represents training and cross validation of the optimized classifier; the green arrow indicates the final classification map generation and accuracy assessment using the trained classifiers and ground truth data.

3.1. Neural Network Classifiers

The Conv1D deals with one-dimensional features and adopts one-dimensional convolution filters. It can be used for hyperspectral data or time series data by capturing temporal features or spectral features of the input data at pixel level [18]. A Conv1D classifier generally contains convolution layers, pooling layers, dense layers, and an output layer. Through applying different convolution filters within each convolution layer, the Conv1D can extract different one-dimensional features from different layers. The pooling layers are generally used for dimension reduction, and the pooling layers are optional for a Conv1D classifier. Dense layers are simple neural network layers. Dropout is a technique of dropping out some neurons in hidden layers randomly during training to prevent overfitting of training data [31]. Therefore, a dropout rate is usually applied to the convolution layers and the dense layers. To build a Conv1D architecture for this study, the VGG16 [32] convolutional neural network is modified for one dimensional data. VGG16 is one of the famous models for 2D image classification. It contains five blocks of convolutional layers and pooling layers combined with three fully connected layers. The first two blocks have 2 padding layers, 2 convolution layers, and 1 pooling layer. The next three blocks have 3 padding layers, 3 convolution layers, and 1 pooling layer respectively. The numbers of convolution filters for each block are 64, 128, 256, 512 and 512, and pooling layers were fixed as max-pooling with a pooling size of 2. One flattened layer is added to the last pooling layer, and two dense layers with 4096 neurons and one output layer are followed. To build the best Conv1D for our study, the 1D-VGG16 was tested by removing 1, 2, 3, or 4 of the blocks, and removing the pooling layers. The convolution filter widths of 3, 5, and 7 were tested and the values of 256, 512, 1024, 2048, 4096 were tested for the number of neurons in the two dense layers. The values of 0, 0.2, 0.5, and 0.8 were also tested for dropout rate for the three fully connected layers.

The MLP is a simple deep feedforward neural network [18], and there are at least three layers (input layer, hidden layer, and output layer) of nodes or neurons. For MLP, the search range of hidden layer number was tested from 1 to 5. The number of neurons in each layer was set to the same, and the

numbers of 64, 128, 512, and 1024 were tested. The values of 0, 0.2, 0.5, and 0.8 were tested for dropout rate. Long short-term memory (LSTM) is a special RNN unit that is capable of learning long-term dependencies or capturing long distance connections in sequence prediction problems [33]. It is composed of memory cells to remember information for long periods of time. For LSTM, the search range of LSTM layer number was tested from 1 to 5. The number of neurons in each layer was set to the same, and the numbers of 64, 128, 256, and 512 were tested. The values of 0, 0.2, 0.5, and 0.8 were tested for dropout rate. Same as Conv1D, three dense layers were added to the last LSTM layer. The values of 256, 512, 1024, 2048, 4096 were tested for the number of neurons in the two dense layers.

The three deep learning architectures were trained using the Adam optimizer [34]. A maximum number of epochs was set to 20 using an early stopping with a patience value of zero. The values of 32, 320, and 3200 were tested for batch size. Parameters of Adam were fixed as: learning rate = 0.00015, β1 = 0.9, β2 = 0.999. As the classes are represented using integers, sparse_categorical_crossentropy loss function was adopted. The three deep learning classification models were built and evaluated using the Keras library [35] on top of Tensorflow [36], and were trained on NVIDIA GeForce RTX 2080Ti Graphical Processing Units (GPUs).

3.2. Other Classifiers

Three efficient machine learning classifiers XGBoost, RF and SVM were used as benchmark classifiers. The XGBoost is a state-of-the-art algorithm that has been growing in popularity in data science due to its accuracy and scalability. It is an implementation of gradient tree boosting technique designed for high efficiency and performance [37]. Zhong et al. [18] tested this algorithm in crop classification using remote sensing time series data.

The RF classifier is an ensemble of decision tree classifiers. Each tree is grown to the maximum depth independently using a random combination of features from the input features [15]. It has been widely used in remote sensing image classifications due to its high performance and less overfitting characteristics.

The SVM classifier can separate hyperplanes and perform non-linear classification using kernel functions [38]. SVM also has been extensively applied in remote sensing classification tasks [18,39,40]. The hyperparameters of the three classifiers were selected after running a grid search on the training dataset. The hyperparameters we tested and adopted in this study are shown in Table 3.

Table 3. Parameter values of the three non-deep learning classifiers used in this study.

Classifier	Parameters	Candidate Values	Selected Values
	Learning rate	0.01, 0.015, 0.05, 0.1	0.1
	Gamma	0, 0.05, 0.1, 0.5	0
	Max_depth	3, 5, 7, 9	3
	Min_child_weight	1, 3, 5, 7	1
XGBoost	Colsample_bytree	0.7, 0.8, 0.9, 1	1
	Subsample	0.7, 0.8, 0.9, 1	1
	N_estimators	50, 100, 150, 200	100
	Reg_lamda	0.01, 0.1, 1	1
	Reg_alpha	0, 0.1, 0.5, 1	0
	N_estimators	50, 100, 150, 200	100
	Max_depth	5, 10, 15, 20, None	None
RF	Min_samples_split	2, 5, 10, 100	2
	Min_samples_leaf	1, 2, 5, 10	1
	Max_features	'log2', 'sqrt', None	'sqrt'
	Kernel function	'linear', 'poly', 'rbf'	'linear'
SVM	C	0.1, 1, 10, 100	1
	Gamma	'scale', 'auto', 0.1, 1, 10	'scale'

3.3. Evaluation

To assess the performance of each algorithm, the confusion matrix was generated using the testing dataset. The producer's accuracy (PA), user's accuracy (UA), overall accuracy (OA) and the Kappa coefficient were computed. In addition, F1 score was used as an indicator of classification accuracy for each class [9,18]. F1 score is the harmonic mean of producer's accuracy and user's accuracy.

$$F1 = \frac{1}{\frac{1}{2}\left(\frac{1}{PA} + \frac{1}{UA}\right)} = \frac{2PA \times UA}{PA + UA} \tag{1}$$

4. Results

Table 4 shows the hyper-parameters of the three deep learning methods. The Conv1D performances the best when the last block of VGG16 networks without the pooling layer and the fully connected layers were kept. The number of neurons on the dense layer is 512 instead of 4096, and the dropout rate is 0.5 for the dense layers. The proposed architecture of the Conv1D is shown in Figure 3. The optimized MLP architecture includes 1 input layer, 3 hidden layers and 1 output layer. The input layer and each hidden layer have 512 neurons. The dropout rate is 0.5. The optimized LSTM-based RNN model contains three LSTM units with 256 output channels for each unit and followed by dropout of 0.5. Similarly to the Conv1D, there are 512 neurons for each dense layer and the dropout rate is 0.5.

Table 4. Summary of the hyper-parameters of the optimized deep learning classifiers.

Classifier	Parameters		Values
Conv1D	Convolution Layers	2, 3, 6, 8, 10	3
	Learning rate	0.01, 0.001, 1.5×10^{-5}, 1.5×10^{-4}, 1.0×10^{-4}	1.5×10^{-4}
	Epoch	10, 20, 30	20
	Batch size	32, 320, 3200	3200
	Dropout	0, 0.2, 0.5, 0.8	0.5
MLP	Layers	1, 2, 3, 4, 5, 6	5
	Nodes each layer	64, 128, 256, 512, 1024	512
	Learning rate	0.01, 0.001, 1.5×10^{-5}, 1.5×10^{-4}, 1.0×10^{-4}	1.5×10^{-4}
	Epoch	10, 20, 50, 100	20
	Batch size	32, 320, 3200	3200
	Dropout	0, 0.2, 0.5, 0.8	0.5
LSTM	Layers	1, 2, 3, 4	3
	Nodes each layer	32, 64, 128, 256, 512	256
	Epoch	10, 20, 30	10
	Batch size	32, 320, 3200	320
	Learning rate	0.001, 1.0×10^{-4}, 1.5×10^{-4}, 1.5×10^{-5}	1.5×10^{-4}
	Dropout	0, 0.2, 0.5, 0.8	0.5

All the polarimetric SAR parameters (coherency matrix, linear polarizations, Pauli decomposition, Cloude-Pottier decomposition, Freeman-Durden decomposition) described in Section 2.4 were tested by all the classifiers. It was found that the Pauli decomposition gave the best OA among the polarimetric parameters. To explore the potential of coherency matrix, the coherency matrix and Pauli decomposition were combined with the multispectral data respectively. The results of Freeman-Durden decomposition, Cloude-Pottier decomposition, and linear polarizations are not listed here in order to focus on the coherency matrix and the best polarimetric SAR parameters. All the classifiers were run on ten scenarios (Table 5): (1) the coherency matrix of the 10 RADARSAT-2 images; (2) Pauli decomposition; (3) linear polarization; (4) Cloude-Pottier decomposition; (5) Freeman-Durden decomposition; (6) the two VENμS images only; (7) two VENμS images and coherency matrix of RADARSAT-2 images (Coherency matrix+ VENμS); (8) MNF transformation of VENμS and coherency matrix, denoted by MNF (Coherency matrix+ VENμS); (9) two VENμS images and Pauli decomposition; (10) MNF transformation of VENμS images and Pauli decomposition RADARSAT-2 images, denoted by MNF (Pauli+ VENμS). No post-classification process was applied to the results in order to assess the performance of the algorithms directly.
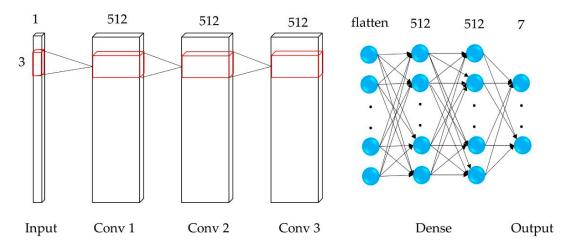
Figure 3. Proposed architecture of Conv 1D. The network input is a feature series of one pixel. Three convolutional layers are consecutively applied, then three dense layers, and the output layer, that provides the predicting class.

Table 5. The input data of different scenarios.

Scenario	Input Data
Scenario 1	Coherency matrix
Scenario 2	Pauli decomposition
Scenario 3	linear polarization
Scenario 4	Cloude-Pottier decomposition
Scenario 5	Freeman-Durden decomposition
Scenario 6	VENμS
Scenario 7	Coherency matrix+ VENμS
Scenario 8	MNF (Coherency matrix+ VENμS)
Scenario 9	Pauli+ VENμS
Scenario 10	MNF (Pauli+ VENμS)

4.1. Overall Classification Accuracy and Training Time

Table 6 presents the average OA (± standard deviation), Kappa coefficient and training time over five folds of training datasets for all methods from Scenario 1 to Scenario 10.

The results show that (1) the Pauli decomposition gave the best OA among other polarimetric SAR parameters. (2) The combination of multi-temporal RADARSAT-2 polarimetric SAR data and VENμS optical data performed better than using either data alone. This is because the multi-temporal RADARSAT-2 polarimetric SAR can capture the structure information of the land covers at temporal scale and the VENμS optical images can provide rich spectral information. The combination of the two sources of data provides not only spectral features, but also structure features and temporal features of the land covers. The MNF transformation further improved the classification accuracies because the noise in the raw data was segregated after MNF transformation. (3) The MNF transformation of Pauli+ VENμS and the MNF transformation of coherency matrix+ VENμS gave similar and the highest OA when Conv1D was applied (96.65 ± 1.03% vs. 96.72 ± 0.77%), which indicates that the MNF transformation is able to extract information from the raw data and the Conv1D has the best learning capability. (4) The Conv1D also performed the best among all the classifiers when coherency matrix only (OA = 91.85 ± 2.51%), or VENμS data only were utilized (OA = 93.15 ± 2.06%). Through MNF transformation, the information and noise were reordered. The first band contains the most information and the last few bands are basically noise. In terms of the efficiency, among the deep learning methods, the MLP needs the least time, which is mostly less than 1 min, to be trained over the five folds of training dataset. While the LSTM needs the most time (13–44 min). The Conv1D needs about 3 min to 11 min. Among the three non-deep learning methods, the RF is the most efficient classifier (2–0 min), while the SVM is the least efficient classifier (20 min—17.5 h).

Table 6. Average OA (±standard deviation), Kappa coefficient, and training time of the six classifiers for different input datasets.

	OA (%)					
	Conv1D	**MLP**	**LSTM**	**XGBoost**	**RF**	**SVM**
Scenario 1	91.85 ± 2.51	91.48 ± 2.38	85.75 ± 1.22	91.65 ± 2.21	91.55 ± 2.38	90.87 ± 2.12
Scenario 2	92.46 ± 2.54	92.47 ± 2.24	87.13 ± 2.02	92.95 ± 2.24	93.05 ± 2.36	91.50 ± 2.31
Scenario 3	90.40 ± 2.88	90.70 ± 2.89	83.05 ± 7.26	90.84 ± 2.93	91.17 ± 3.21	89.54 ± 2.82
Scenario 4	90.36 ± 2.42	86.97 ± 2.16	88.37 ± 2.55	89.01 ± 2.07	89.22 ± 1.89	87.50 ± 1.36
Scenario 5	88.68 ± 2.07	58.60 ± 1.97	90.03 ± 3.79	91.99 ± 2.55	91.59 ± 2.21	89.62 ± 3.21
Scenario 6	93.15 ± 2.06	91.42 ± 3.01	90.02 ± 0.65	92.34 ± 1.88	92.63 ± 2.07	93.23 ± 1.76
Scenario 7	94.71 ± 1.85	94.84 ± 1.39	94.11 ± 0.42	95.23 ± 1.81	94.43 ± 2.26	95.30 ± 0.92
Scenario 8	**96.65 ± 1.03**	95.67 ± 2.17	93.82 ± 0.77	94.82 ± 1.57	95.39 ± 1.71	95.90 ± 0.69
Scenario 9	95.31 ± 1.78	94.58 ± 1.38	92.91 ± 0.59	95.38 ± 1.72	95.09 ± 1.70	95.23 ± 1.14
Scenario 10	**96.72 ± 0.77**	96.14 ± 1.47	93.65 ± 2.47	95.45 ± 1.29	95.92 ± 1.12	96.06 ± 0.86
	Kappa					
	Conv1D	**MLP**	**LSTM**	**XGBoost**	**RF**	**SVM**
Scenario 1	0.87 ± 0.04	0.87 ± 0.04	0.79 ± 0.12	0.87 ± 0.03	0.87 ± 0.04	0.86 ± 0.03
Scenario 2	0.89 ± 0.04	0.89 ± 0.03	0.81 ± 0.03	0.89 ± 0.03	0.90 ± 0.04	0.87 ± 0.03
Scenario 3	0.86 ± 0.04	0.86 ± 0.04	0.74 ± 0.11	0.86 ± 0.04	0.87 ± 0.05	0.84 ± 0.04
Scenario 4	0.85 ± 0.04	0.80 ± 0.03	0.83 ± 0.04	0.83 ± 0.03	0.84 ± 0.03	0.81 ± 0.02
Scenario 5	0.83 ± 0.03	0.31 ± 0.03	0.85 ± 0.06	0.88 ± 0.04	0.87 ± 0.03	0.84 ± 0.05
Scenario 6	0.90 ± 0.03	0.87 ± 0.05	0.85 ± 0.01	0.88 ± 0.03	0.89 ± 0.03	0.90 ± 0.03
Scenario 7	0.92 ± 0.03	0.92 ± 0.02	0.91 ± 0.01	0.93 ± 0.03	0.92 ± 0.03	0.93 ± 0.01
Scenario 8	**0.95 ± 0.02**	0.93 ± 0.03	0.91 ± 0.01	0.92 ± 0.02	0.93 ± 0.03	0.94 ± 0.01
Scenario 9	0.93 ± 0.03	0.92 ± 0.02	0.89 ± 0.01	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.02
Scenario 10	**0.95 ± 0.01**	0.94 ± 0.02	0.90 ± 0.04	0.93 ± 0.02	0.94 ± 0.02	0.94 ± 0.01
	Training time (s)					
	Conv1D	**MLP**	**LSTM**	**XGBoost**	**RF**	**SVM**
Scenario 1	707.35	51.00	2160.15	3588.45	1231.10	3977.45
Scenario 2	269.70	50.50	876.10	1204.55	561.20	1499.00
Scenario 3	238.20	49.25	884.00	1204.10	555.35	2074.30
Scenario 4	192.60	61.45	877.25	1224.65	563.90	3628.15
Scenario 5	249.65	37.45	888.00	1129.25	467.65	1837.10
Scenario 6	281.65	57.55	787.10	530.30	127.25	1584.85
Scenario 7	657.20	50.80	2666.50	3764.60	969.20	1838.75
Scenario 8	348.55	47.50	2675.55	4352.60	1007.20	63,082.15
Scenario 9	345.80	45.35	1408.05	1638.05	526.15	1190.90
Scenario 10	175.50	40.25	1401.65	2075.60	689.50	21,200.10

4.2. Classification Accuracy of Individual Land Cover Class

The F-score of each class was calculated for different classifiers and different input datasets (Scenario 1–10) according to Equation (1) (Table 7). The highest F-score values for Soybean (97.23 ± 0.66%), Corn (97.60 ± 0.87%), and Other (86.37 ± 9.14%) were generated by the Conv1D, and the highest F-score values for Wheat (98.69 ± 0.92%), Forage (94.06 ± 1.97%), and Built-up (99.79 ± 0.22%) were given by the LSTM. The RF gave the best accuracy on Forest (98.91 ± 1.15%). However, the accuracies of individual classes vary a lot when the input datasets vary. The performance of LSTM seems not as stable as the Conv1D. The F-score values of Soybean, Corn, Forest and Built-up are higher than 90% except for the LSTM classifier when the optical data were utilized. The class Other has the worst classification accuracy among all the classes due to the limited number of training samples. The combination of multi-temporal RADARSAT-2 and VENμS data improved the accuracies of all classes but Built-up for all the classifiers except the LSTM. The MNF transformation improved the accuracies of most classes but forest when the Conv1D was applied, and the class Other was improved the most. From Scenario 7 to Scenario 10, the F-score values of Soybean and Corn, which are two of the three main crops in this study site, are similar no matter the coherency matrix or the Pauli decomposition were utilized. The confusion matrix of the best overall accuracy result is shown in Table 8. Misclassifications mainly occur between corn and soybean, wheat and forage, and soybean and other.

Table 7. Averaged F-score (± standard deviation) of each class for the six classifiers. The best F-score of each class was highlighted in bold.

	Input	Soybean	Corn	Wheat	Forage	Forest	Built-Up	Other
Conv1D	Scenario 1	93.50 ± 2.29	94.87 ± 1.74	86.17 ± 5.26	73.91 ± 9.71	86.83 ± 3.48	87.65 ± 13.48	47.93 ± 15.13
	Scenario 2	93.99 ± 2.17	94.51 ± 2.15	89.00 ± 6.36	75.19 ± 10.89	93.04 ± 2.25	85.21 ± 15.11	63.36 ± 12.60
	Scenario 3	92.07 ± 2.67	93.07 ± 2.76	85.31 ± 6.21	70.46 ± 11.24	87.76 ± 11.24	87.32 ± 1.16	53.90 ± 20.18
	Scenario 4	90.51 ± 1.38	94.57 ± 1.88	80.38 ± 8.36	71.06 ± 9.27	63.83 ± 6.22	62.55 ± 30.37	0
	Scenario 5	92.14 ± 2.18	93.03 ± 2.08	88.09 ± 6.26	69.40 ± 8.34	87.74 ± 3.29	79.32 ± 13.98	28.44 ± 10.37
	Scenario 6	93.09 ± 1.84	95.87 ± 1.85	92.07 ± 9.63	77.72 ± 13.22	97.31 ± 1.22	98.91 ± 0.75	31.77 ± 18.93
	Scenario 7	95.12 ± 2.12	96.25 ± 1.17	94.97 ± 3.22	83.28 ± 9.22	97.93 ± 1.38	91.48 ± 13.64	54.30 ± 17.22
	Scenario 8	**97.23 ± 0.66**	**97.60 ± 0.87**	95.34 ± 5.00	87.89 ± 8.75	97.43 ± 2.72	98.49 ± 1.38	81.14 ± 9.60
	Scenario 9	95.39 ± 1.97	96.60 ± 1.42	95.11 ± 5.08	86.92 ± 7.57	98.23 ± 1.24	93.54 ± 11.17	59.53 ± 16.17
	Scenario 10	97.13 ± 0.72	97.51 ± 0.97	96.07 ± 4.38	88.56 ± 7.95	98.21 ± 2.08	96.10 ± 4.91	**86.37 ± 9.14**
MLP	Scenario 1	93.26 ± 1.93	94.90 ± 1.39	84.66 ± 9.05	73.53 ± 10.96	86.28 ± 2.99	85.43 ± 13.92	2.74 ± 4.73
	Scenario 2	93.71 ± 1.87	94.99 ± 1.70	89.55 ± 6.40	77.03 ± 10.49	92.10 ± 3.03	82.92 ± 16.17	27.31 ± 24.16
	Scenario 3	92.12 ± 2.50	93.51 ± 2.68	87.09 ± 5.96	72.58 ± 10.23	87.82 ± 1.29	85.47 ± 12.79	19.75 ± 25.85
	Scenario 4	88.79 ± 1.82	94.17 ± 2.19	81.82 ± 6.75	68.07 ± 9.59	40.42 ± 7.47	45.00 ± 27.11	0
	Scenario 5	62.65 ± 1.60	67.20 ± 3.59	14.83 ± 7.05	2.34 ± 1.29	1.40 ± 0.94	0	0
	Scenario 6	91.68 ± 3.07	94.65 ± 2.15	88.63 ± 8.71	70.03 ± 11.88	95.18 ± 2.86	99.20 ± 0.8	0
	Scenario 7	95.59 ± 1.80	96.37 ± 0.91	95.57 ± 3.43	86.39 ± 8.24	98.05 ± 1.24	91.11 ± 10.83	25.19 ± 24.91
	Scenario 8	96.11 ± 1.91	97.18 ± 1.23	94.47 ± 6.37	86.00 ± 11.12	97.00 ± 3.10	94.71 ± 4.33	52.26 ± 23.31
	Scenario 9	94.81 ± 1.66	96.42 ± 1.23	94.62 ± 3.67	85.08 ± 6.57	98.00 ± 0.93	92.12 ± 13.26	20.13 ± 13.66
	Scenario 10	96.54 ± 1.49	97.35 ± 1.07	95.69 ± 4.39	87.35 ± 8.58	97.48 ± 2.05	95.73 ± 3.86	71.79 ± 10.71
LSTM	Scenario 1	87.64 ± 1.63	95.46 ± 0.20	69.94 ± 1.59	71.91 ± 1.95	70.57 ± 1.94	77.99 ± 14.79	17.82 ± 3.55
	Scenario 2	88.58 ± 2.76	95.12 ± 0.27	74.28 ± 1.26	70.10 ± 1.23	80.06 ± 1.92	91.78 ± 3.89	23.43 ± 8.12
	Scenario 3	87.82 ± 5.89	88.45 ± 7.63	62.42 ± 15.57	45.19 ± 18.08	76.27 ± 5.61	69.90 ± 21.65	47.46 ± 18.36
	Scenario 4	89.32 ± 2.65	94.40 ± 1.67	83.54 ± 10.72	75.80 ± 5.47	66.96 ± 5.45	76.11 ± 13.96	40.05 ± 18.78
	Scenario 5	91.77 ± 3.56	92.52 ± 4.32	84.89 ± 9.49	72.24 ± 5.48	88.40 ± 1.89	79.08 ± 18.31	66.67 ± 13.46
	Scenario 6	87.81 ± 0.82	90.58 ± 0.35	96.80 ± 0.95	88.31 ± 2.44	96.75 ± 1.02	98.43 ± 0.08	23.44 ± 12.99
	Scenario 7	95.70 ± 0.25	96.21 ± 0.26	93.21 ± 0.98	66.09 ± 9.67	93.56 ± 2.54	**99.79 ± 0.22**	49.76 ± 5.98
	Scenario 8	95.20 ± 0.47	97.48 ± 0.65	96.92 ± 1.55	57.80 ± 3.91	90.39 ± 1.36	86.08 ± 5.71	32.15 ± 4.12
	Scenario 9	92.50 ± 0.79	94.79 ± 0.41	97.31 ± 0.64	87.21 ± 3.21	98.32 ± 0.29	89.54 ± 10.92	43.10 ± 13.12
	Scenario 10	92.81 ± 3.41	94.47 ± 2.53	**98.69 ± 0.92**	**94.06 ± 1.97**	92.86 ± 2.40	92.55 ± 2.35	54.72 ± 3.95
XGBoost	Scenario 1	93.29 ± 1.99	94.64 ± 1.88	86.94 ± 8.70	71.52 ± 2.11	86.73 ± 2.11	87.12 ± 18.89	57.79 ± 9.46
	Scenario 2	94.19 ± 1.71	95.03 ± 1.95	90.26 ± 5.18	76.91 ± 9.51	91.65 ± 1.90	85.53 ± 19.40	65.56 ± 11.04
	Scenario 3	92.53 ± 2.63	93.61 ± 2.74	86.45 ± 6.12	70.64 ± 9.46	87.07 ± 2.11	86.19 ± 18.76	58.99 ± 11.80
	Scenario 4	90.47 ± 1.13	94.48 ± 1.93	82.90 ± 10.90	71.78 ± 12.71	61.52 ± 6.85	67.25 ± 31.69	40.27 ± 9.44
	Scenario 5	93.33 ± 1.93	94.05 ± 1.94	90.84 ± 8.07	72.78 ± 7.08	89.95 ± 1.36	84.20 ± 17.79	60.98 ± 10.49
	Scenario 6	92.16 ± 2.00	94.44 ± 1.62	90.75 ± 8.73	77.68 ± 10.07	97.36 ± 1.95	98.58 ± 1.54	69.15 ± 5.54
	Scenario 7	95.44 ± 1.72	96.29 ± 1.48	95.27 ± 5.06	85.54 ± 9.5	98.39 ± 1.40	93.61 ± 9.84	72.00 ± 13.98
	Scenario 8	95.61 ± 0.78	96.84 ± 0.89	91.36 ± 8.38	81.31 ± 11.16	96.57 ± 2.44	90.46 ± 16.55	69.32 ± 17.62
	Scenario 9	95.67 ± 1.49	96.40 ± 1.48	94.97 ± 5.59	85.46 ± 9.16	98.65 ± 1.42	94.37 ± 9.94	75.99 ± 11.46
	Scenario 10	95.75 ± 1.14	96.93 ± 0.95	94.93 ± 3.20	83.29 ± 9.10	96.84 ± 2.13	93.24 ± 12.01	79.12 ± 10.18
RF	Scenario 1	92.74 ± 2.02	94.48 ± 2.08	87.49 ± 6.70	73.19 ± 8.83	86.10 ± 2.76	89.69 ± 18.07	59.82 ± 13.94
	Scenario 2	94.14 ± 1.84	95.02 ± 2.00	90.39 ± 6.98	77.79 ± 10.37	91.55 ± 2.10	89.40 ± 17.79	71.92 ± 11.25
	Scenario 3	92.47 ± 2.86	93.68 ± 2.96	87.54 ± 7.30	72.68 ± 9.83	86.98 ± 2.17	90.03 ± 16.95	70.53 ± 16.17
	Scenario 4	90.46 ± 1.17	94.62 ± 2.01	83.47 ± 10.89	73.74 ± 9.64	60.52 ± 6.05	65.36 ± 33.79	37.63 ± 9.76
	Scenario 5	93.16 ± 1.81	93.72 ± 2.36	88.19 ± 7.20	72.17 ± 7.12	89.28 ± 1.51	89.51 ± 18.19	63.54 ± 11.08
	Scenario 6	92.82 ± 2.08	94.41 ± 1.80	93.09 ± 8.43	72.81 ± 15.67	97.64 ± 1.76	99.26 ± 0.52	77.56 ± 11.71
	Scenario 7	94.59 ± 2.09	95.52 ± 1.53	94.99 ± 6.21	81.82 ± 12.26	98.64 ± 1.22	96.87 ± 5.86	79.35 ± 10.13
	Scenario 8	95.48 ± 1.29	96.95 ± 0.78	93.27 ± 7.77	85.08 ± 10.10	97.87 ± 2.19	95.98 ± 6.91	64.03 ± 11.15
	Scenario 9	95.17 ± 1.83	96.17 ± 1.21	95.93 ± 3.78	83.37 ± 11.50	**98.91 ± 1.15**	96.00 ± 7.27	81.97 ± 9.39
	Scenario 10	96.10 ± 0.92	97.17 ± 0.73	95.76 ± 3.61	85.36 ± 10.14	97.98 ± 1.94	93.36 ± 11.56	79.06 ± 10.15
SVM	Scenario 1	92.02 ± 2.29	94.34 ± 1.65	87.55 ± 5.39	75.21 ± 9.01	87.52 ± 3.95	87.17 ± 16.15	40.96 ± 18.71
	Scenario 2	92.38 ± 2.50	94.65 ± 1.73	90.13 ± 5.65	76.41 ± 10.24	92.12 ± 2.54	84.90 ± 17.21	47.03 ± 19.76
	Scenario 3	90.64 ± 2.68	93.25 ± 2.37	86.45 ± 6.97	72.40 ± 9.57	87.37 ± 3.11	85.05 ± 17.20	35.51 ± 16.05
	Scenario 4	89.01 ± 1.84	94.30 ± 2.06	74.69 ± 9.83	71.02 ± 11.01	56.84 ± 6.20	67.08 ± 35.41	5.81 ± 11.61
	Scenario 5	90.99 ± 3.27	93.00 ± 2.79	86.82 ± 6.69	71.57 ± 8.77	90.12 ± 0.97	83.87 ± 18.20	45.90 ± 19.81
	Scenario 6	92.91 ± 1.70	95.80 ± 1.60	90.85 ± 8.51	79.91 ± 11.64	97.41 ± 0.89	99.38 ± 0.55	36.51 ± 17.09
	Scenario 7	95.68 ± 1.47	96.82 ± 1.06	95.45 ± 3.67	87.21 ± 7.52	98.51 ± 1.33	97.54 ± 4.55	56.38 ± 20.15
	Scenario 8	96.63 ± 0.76	97.31 ± 0.78	95.03 ± 3.47	82.92 ± 8.92	97.61 ± 1.97	94.59 ± 10.36	77.45 ± 19.18
	Scenario 9	95.34 ± 1.84	96.98 ± 1.34	96.12 ± 3.72	87.57 ± 8.01	98.31 ± 1.65	96.39 ± 5.77	60.04 ± 21.40
	Scenario 10	96.67 ± 0.92	97.37 ± 0.99	95.62 ± 3.71	83.76 ± 9.38	98.48 ± 1.26	93.04 ± 12.17	82.03 ± 20.87

Table 8. Confusion matrix of Scenario 10 for the Conv1D.

Class	Ground Truth						
	Soybean	Corn	Wheat	Forage	Forest	Built-Up	Other
Soybean	19818	55	6	71	22	0	28
Corn	778	21052	0	31	0	25	114
Wheat	0	0	3789	11	0	0	0
Forage	4	0	82	2914	0	0	0

Table 8. *Cont.*

Class	Ground Truth						
	Soybean	Corn	Wheat	Forage	Forest	Built-Up	Other
Forest	1	0	0	56	2443	0	0
Built-up	0	0	0	0	0	480	0
Other	42	0	0	0	0	0	328

All the classifiers were applied to the MNF transformation of Pauli decomposition and VENµS data (Scenario 10) to make classification for the whole study area. The classification maps generated by the four classifiers Conv1D, MLP, LSTM, SVM, XGBoost, RF are shown in Figure 4a–f respectively. Three major differences are marked using black boxes. The MLP shows more misclassifications between Other and Soybean than the Conv1D as the small cyan patches spread in the purple soybean fields. The non-deep learning methods show more misclassification between winter wheat and forage. This is reasonable because the forage and winter wheat started to green up at the same time of the growing season. The Conv1D also gave misclassification between these classes. However, the accuracy of the marked areas cannot be validated since there are no accurate reference data in those areas.
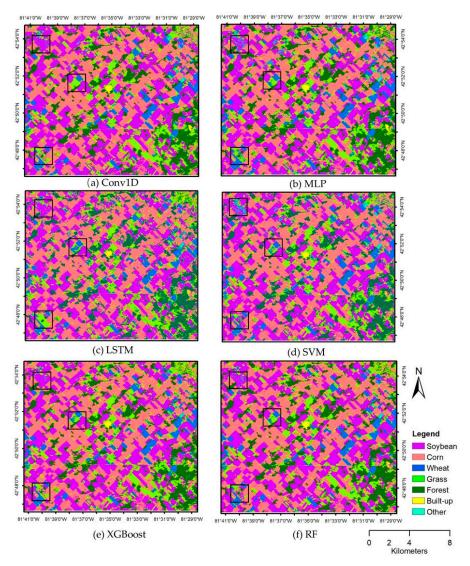
Figure 4. Classification maps of the four classifiers (**a**) Conv1D, (**b**) MLP, (**c**) LSTM, (**d**) SVM, (**e**) XGBoost, (**f**) RF using the MNF transformation of RADARSAT-2 Pauli decomposition and VENµS data. Misclassifications mainly exist between Wheat and Forage, as well as Soybean and Other.

5. Discussions

In this study, multi-temporal RADARSAT-2 polarimetric SAR data and VENμS data were acquired. It is worth noting that the RADARSAT-2 constellation was launched and multi-temporal full polarimetric SAR data will be available for free to users (subject to security restrictions set out in Canadian legislation). The VENμS data are currently only available at a few specific areas in the world. The data were utilized for the first time and they show good potential in cropland classification. Sentinel-2 data can also be used instead if this approach will be applied in other study areas.

5.1. Comparisons between Conv1D and Other Classifiers

In this study, we found that the use of pooling layer will lower the performance of the Conv1D, which confirmed the findings of [9]. For example, when one pooling layer was added to the Conv1D architecture, the average overall accuracy was 96.06 ± 1.53% for Scenario 10, while the average overall accuracy of Conv1D without the pooling layer was 96.72 ± 0.77%. The Conv1D gave the highest overall classification accuracy when the MNF transformation of multi-temporal RADARSAT-2 (both coherency matrix and Pauli decomposition) and the VENμS data were utilized. In terms of the execution time, the MLP is the most efficient classifier and the Conv1D is more efficient than the LSTM and the three non-deep learning methods. The LSTM-based RNN performs the worst among the deep learning and non-deep learning methods and it needs the more time to be trained than other two deep learning methods and the RF classifier. Therefore, the LSTM seems not suitable for the datasets in this specific classification task. The SVM also performs well for the combination of multi-temporal RADARSAT-2 and VENμS data, but it needs the most time to be trained among all the classifiers. A paired *t*-test was performed to compare the mean OA of Conv1D against the other classifiers when the MNF transformation of multi-temporal RADARSAT-2 and VENμS data. The *p*-value is about 0.1 for Conv1D and MLP, which indicates that the Conv1D outperforms the MLP but not significantly. However, the MLP performs significantly worse than the Conv1D on the Cloude-Pottier decomposition, Freeman-Durden decomposition and the two VENμS data. The *p*-values are lower than 0.05 for Conv1D and other classifiers, which means that the Con1D significantly outperforms the LSTM, XGBoost, RF and SVM. These indicate that the Conv1D has a great potential in classification tasks using the combination of multi-spectral, multi-temporal, and multi-modal data.

5.2. Influence of Input Features on the Performance of Conv1D

This study shows that the performances of the deep learning methods still depend on the feature selection. The coherency matrix contains all the raw information of the polarimetric SAR data, but it also contains noise, which will affect the performance of the classifiers. The two multi-spectral data can produce higher classification accuracy than RADARSAT-2 data. This may be because the speckle noise in RADARSAT-2 data and there is a lack of RADARSAT-2 data in May and June. The combination of RADARSAT-2 and VENμS data significantly improved the accuracy for Grass, Forest, Built-up, and Other class, and slightly improved the accuracy of soybean.

Among the four polarimetric SAR parameters, the overall classification accuracy of Pauli decomposition is the highest, which confirms the finding of a previous study [15]. The components of the Pauli decomposition can be represented by the three diagonal elements of the coherency matrix, which represent single bounce scattering (e.g., bare soil), double bounce scattering (e.g., soil-stalk), and volume scattering (e.g., crop canopies), respectively [41]. While the linear polarizations can be represented by the three diagonal elements of the covariance matrix. Freeman-Durden decomposition decomposes each polarimetric SAR backscatter three components that are modeled as the first-order Bragg surface scatter, the scattering from a dihedral corner reflector, and canopy scatters from randomly oriented dipoles respectively [42]. The Cloude-Pottier decomposition decomposes the coherency matrix into entropy (H), anisotropy (A), and alpha angle (α) [43]. These parameters reflect the crop structural features.

When directly stacking the multi-temporal RADARSAT-2 coherency matrix elements and VENµS multi-spectral data, the performances of the three deep learning methods are similar or even inferior to the three classical machine learning methods. However, when MNF transformation was applied to the original data, the overall classification accuracies improved about 1% for the Conv1D and MLP. For Conv1D, the OA of the two MNF transformation data are similar, indicating that the MNF transformation of the raw data can extract information as much as the MNF transformation of the best feature manually selected. As we know, the manual selection of features is time consuming, and we cannot guarantee that the feature we arbitrarily selected is the best.

In this study, we compared the MNF transformation and the PCA transformation, we found that the results of MNF (96.65%) is superior to PCA (95.36%) when Conv1D was applied. It demonstrated that the MNF transformation can also be used as a multi-source data fusion method in land use land cover (LULC) classification applications.

5.3. Extended Experiments on Datasets with Different Number of RADARSAT-2 and VENµS Data

To test the robustness of the Conv1D classifier and the contribution of the MNF transformation and the coherency matrix, the proposed architecture was tested on a dataset with different number of RADARSAT-2 and VENµS data. And this is the case that we cannot acquired as many dates of data as we used in this study site, especially the RADARSAT-2 data. We selected the RADARSAT-2 acquired on 1 July, 25 July, 18 August, 1 September, 15 September, and VENµS data acquired on 9 July as Sub-dataset 1 and the RADARSAT-2 acquired on 8 July, 1 August, 25 August, 8 September, 5 October, and VENµS data acquired on 11 June as Sub-dataset 2. The results (Table 9) show that the combination of Pauli decomposition and VENµS multispectral data gave slightly better overall classification accuracy than the combination of coherency matrix and VENµS data for the two sub-datasets (93.86% vs. 93.58% and 92.00% vs. 91.59%). However, the MNF transformation of the two datasets gave the similar overall accuracy (95.85% vs. 95.85% and 92.77% vs. 92.70%). The results confirm the findings that the coherency matrix has a great potential in crop classification and the MNF transformation of multi-temporal RADARSAT-2 and VENµS data can improve the classification accuracy when the Conv1D was adopted. In addition, this experiment indicates that the OA and the increase of OA by MNF transformation depend on the dates of acquired remote sensing data. If there is only one date of optical data, the one acquired on the time when all the crops are present (e.g., middle growing season) with more separable spectral features would be better than the beginning or later in the growing season. In this experiment, the VENµS data acquired on 9 July are better than the ones acquired on 11 June.

Table 9. Average OA (± standard deviation) of the Conv1D using a dataset with a different number of remote sensing data.

	Input	OA (%)	Kappa
Sub-dataset 1	Coherency matrix+ VENµS	93.58 ± 1.55	0.90 ± 0.02
	MNF (Coherency matrix+ VENµS)	95.85 ± 0.95	0.94 ± 0.01
	Pauli+ VENµS	93.86 ± 1.52	0.91 ± 0.02
	MNF (Pauli+ VENµS)	95.85 ± 0.84	0.94 ± 0.01
Sub-dataset 2	Coherency matrix+ VENµS	91.59 ± 3.86	0.87 ± 0.06
	MNF (Coherency matrix+ VENµS)	92.77 ± 2.65	0.89 ± 0.04
	Pauli+ VENµS	92.00 ± 2.98	0.88 ± 0.04
	MNF (Pauli+ VENµS)	92.70 ± 2.93	0.89 ± 0.04

5.4. Future Work

This study shows that among the pixel-based classification methods, the proposed Conv1D performs the best on the MNF transformation of the multi-temporal RADARSAT-2 and VENµS data. However, the proposed Conv1D may not perform well when the number of input data is higher than the number of data that were used for training the hyperparameters. Obviously, the Conv1D does not

consider the spatial features. The 3D CNNs were proven as a more effective deep learning architecture than 1D CNN in crop classification using multi-temporal optical images [18]. The 3D CNNs considered not only the spatial features but also the temporal features. Future work could focus on considering the 3D CNNs on the combination of multi-temporal polarimetric SAR and optical data.

6. Conclusions

In this study, a one-dimensional convolution neural network (Conv1D) was proposed for cropland classification using multi-temporal fully polarimetric SAR and optical data. It was compared with two another deep learning methods including the MLP, and LSTM-based RNN and three non-deep learning methods including the XGBoost, RF, and SVM. We also evaluated the influence of different input features on the performance of the Conv1D by comparing with the benchmark methods. The results show that the performance of all the methods varies with the input datasets. The Conv1D not always performs better than other methods. When VENμS data were directly combined with the coherency matrix, the Conv1D performs slightly inferior to the MLP, XGBoost, and SVM. However, the MNF transformation gave the best OA and improved the F-score values for most classes when the Conv1D was applied. In addition, the MNF transformation of the combination of multi-temporal RADARSAT-2 coherency matrix and VENμS spectral data can give similar OA as the MNF transformation of the combination of Pauli decomposition and VENμS data. These findings indicate that the coherency matrix has great potential in crop classification and the Conv1D can learn features from MNF transformation of multi-temporal RADARSAT-2 and VENμS data better than other classifiers.

References

1. Sonobe, R.; Tani, H.; Wang, X.; Kobayashi, N.; Shimamura, H. Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data. *Remote Sens. Lett.* **2014**, *5*, 157–164. [CrossRef]
2. Huang, X.; Wang, J.; Shang, J.; Liao, C.; Liu, J. Application of polarization signature to land cover scattering mechanism analysis and classification using multi-temporal C-band polarimetric RADARSAT-2 imagery. *Remote Sens. Environ.* **2017**, *193*, 11–28. [CrossRef]
3. Xie, Q.; Wang, J.; Liao, C.; Shang, J.; Lopez-sanchez, J.M. On the use of Neumann decomposition for crop classification using multi-temporal RADARSAT-2 polarimetric SAR data. *Remote Sens.* **2019**, *11*, 776. [CrossRef]
4. Skriver, H. Crop classification by multitemporal C- and L-band single- and dual-polarization and fully polarimetric SAR. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2138–2149. [CrossRef]
5. McNairn, H.; Champagne, C.; Shang, J.; Holmstrom, D.; Reichert, G. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 434–449. [CrossRef]
6. Forkuor, G.; Conrad, C.; Thiel, M.; Ullmann, T.; Zoungrana, E. Integration of optical and synthetic aperture radar imagery for improving crop mapping in northwestern Benin, West Africa. *Remote Sens.* **2014**, *6*, 6472–6499. [CrossRef]

7. Inglada, J.; Vincent, A.; Arias, M.; Marais-Sicre, C. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. *Remote Sens.* **2016**, *8*, 362. [CrossRef]

8. Zhou, T.; Pan, J.; Zhang, P.; Wei, S.; Han, T. Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region. *Sensors* **2017**, *17*, 1210. [CrossRef]

9. Pelletier, C.; Webb, G.I. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens.* **2019**, *11*, 523. [CrossRef]

10. Joshi, N.; Baumann, M.; Ehammer, A.; Fensholt, R.; Grogan, K.; Hostert, P.; Jepsen, M.R.; Kuemmerle, T.; Meyfroidt, P.; Mitchard, E.T.A.; et al. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sens.* **2016**, *8*, 70. [CrossRef]

11. Sukawattanavijit, C.; Chen, J. Fusion of multi-frequency sar data with thaichote optical imagery for maize classification in thailand. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) IEEE, Milan, Italy, 26–31 July 2015; pp. 617–620.

12. Manaf, S.A. Fusion of optical and SAR in extracting shoreline at northeast coast of peninsular Malaysia. In Proceedings of the 36th Asian Conference on Remote Sensing: Fostering Resilient Growth in Asia, ACRS, Quezon City, Philippines, 19–23 October 2015.

13. Basuki, T.M.; Skidmore, A.K.; Hussin, Y.A.; Van Duren, I.; Mutiara, T.; Skidmore, A.K.; Hussin, Y.A.; Van, I. Estimating tropical forest biomass more accurately by integrating ALOS PALSAR and Landsat-7 ETM + data. *Int. J. Remote Sens.* **2013**, *34*, 4871–4888. [CrossRef]

14. Mahyoub, S.; Fadil, A.; Mansour, E.M.; Rhinane, H.; Al-nahmi, F. Fusing of optical and Synthetic Aperture Radar (SAR) remote sensing data: A systematic literature review. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Casablanca, Morocco, 10–11 October 2018; Volume XLII, pp. 127–138.

15. Liao, C.; Wang, J.; Huang, X.; Shang, J. Contribution of minimum noise fraction transformation of multi-temporal radarsat-2 polarimetric sar data to cropland classification. *Can. J. Remote Sens.* **2018**, *44*, 1–17. [CrossRef]

16. Green, A.A.; Berman, M.; Switzer, P.; Graig, M.D. A transformation for ordering multispectral data in term of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 65–74. [CrossRef]

17. Ndikumana, E.; Ho, D.; Minh, T.; Baghdadi, N.; Courault, D.; Hossard, L. Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sens.* **2018**, *10*, 1217. [CrossRef]

18. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [CrossRef]

19. Sun, Z.; Di, L.; Fang, H. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *Int. J. Remote Sens.* **2019**, *40*, 593–614. [CrossRef]

20. Minh, D.H.T.; Ienco, D.; Gaetano, R.; Lalande, N.; Ndikumana, E.; Osman, F.; Maurel, P. Deep Recurrent Neural Networks for Winter Vegetation Quality Mapping via Multitemporal SAR Sentinel-1. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 464–468. [CrossRef]

21. Hu, Y.; Zhang, Q.; Zhang, Y.; Yan, H. A ddeep convolution neural network method for land cover mapping: A case study of Qinhuangdao, China. *Remote Sens.* **2018**, *10*, 2053. [CrossRef]

22. Lim, K.; Jin, D.; Kim, C. Change detection in high resolution satellite images using an ensemble of convolutional neural networks. In Proceedings of the APSIPA Annual Summit and Conference 2018, Honolulu, HI, USA, 12–15 November 2018; pp. 509–515.

23. Ji, S.; Wei, S.; Lu, M. A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. *Int. J. Remote Sens.* **2019**, *40*, 3308–3322. [CrossRef]

24. Audebert, N.; Le Saux, B.; Lefevre, S. Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images. *Remote Sens.* **2017**, *9*, 368. [CrossRef]

25. Qin, Y.; Wu, Y.; Li, B.; Gao, S.; Liu, M.; Zhan, Y. Semantic segmentation of building roof in dense urban environment with deep convolutional neural network: A case study using GF2 VHR imagery in China. *Sensors* **2019**, *19*, 1164. [CrossRef] [PubMed]

26. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sens.* **2018**, *10*, 75. [CrossRef]

27. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]

28. Zhao, H.; Chen, Z.; Jiang, H.; Jing, W.; Sun, L.; Feng, M. Evaluation of three deep learning models for early crop classification using Sentinel-1A imagery time series—A case study in Zhanjiang, China. *Remote Sens.* **2019**, *11*, 2673. [CrossRef]

29. Jiao, X.; Kovacs, J.M.; Shang, J.; McNairn, H.; Walters, D.; Ma, B.; Geng, X. Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 38–46. [CrossRef]

30. Huang, X.; Liao, C.; Xing, M.; Ziniti, B.; Wang, J.; Shang, J.; Liu, J.; Dong, T.; Xie, Q.; Torbick, N. A multi-temporal binary-tree classification using polarimetric RADARSAT-2 imagery. *Remote Sens. Environ.* **2019**, *235*, 111478. [CrossRef]

31. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

32. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Available online: http://adsabs.harvard.edu/abs/2014arXiv1409.1556S (accessed on 19 June 2019).

33. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

34. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015; pp. 1–15.

35. Keras: The Python Deep Learning Library. Available online: https://keras.io/ (accessed on 2 April 2019).

36. Tensorflow: An Open Source Software Library for High Performance Numerical Computation. Available online: https://www.tensorflow.org (accessed on 2 April 2019).

37. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

38. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

39. Lawrence, R.L.; Wood, S.D.; Sheley, R.L. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (Random Forest). *Remote Sens. Environ.* **2006**, *100*, 356–362. [CrossRef]

40. Na, X.; Zhang, S.; Li, X.; Yu, H.; Liu, C. Improved land cover mapping using random forests combined with Landsat Thematic Mapper imagery and ancillary geographic data. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 833–840. [CrossRef]

41. Lee, J.-S.; Pottier, E. *Polarimetric Radar Imaging: From Basics to Applications*; CRC Press: Boca Raton, FL, USA, 2009; ISBN 9781420054972.

42. Freeman, A.; Durden, S.L. A three-component scattering model for polarimetric SAR data. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 963–973. [CrossRef]

43. Cloude, S.R.; Pottier, E. A review of target decomposition theorems in radar polarimetry. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 498–518. [CrossRef]